

# Low-rank approximation of elliptic boundary value problems with high-contrast coefficients\*

M. Bebendorf<sup>†</sup>

October 15, 2014

We analyze the convergence of degenerate approximations to Green's function of elliptic boundary value problems with high-contrast coefficients. It is shown that the convergence is independent of the contrast if the error is measured with respect to suitable norms. This lays ground to fast methods (so-called hierarchical matrix approximations) which do not have to be adapted to the coefficients.

*Keywords:* high-contrast coefficients, hierarchical matrix approximation

## 1 Introduction

Elliptic problems with non-smooth, high-contrast coefficients appear in many fields of science ranging from the simulation of porous media and composite materials to the recent field of uncertainty quantification. The numerical solution of such problems is challenging if all details of the physical problems are to be resolved due to the large number of degrees of freedom needed for a sufficiently accurate discretization. The enormous amount of computer memory and CPU time can be reduced to some extent if one is satisfied with macroscopic properties of the solution. The *multiscale finite element method* [26, 12] and the *heterogeneous multiscale method* [11] capture small-scale effects on large scales. These methods rely on special assumptions on the coefficient such as self-similarities, periodicity and scale separation. If such properties cannot be exploited, then the discretization has to be done with full detail. In this case, the numerical method used to solve the problem has to be efficient and robust with respect to the operator's coefficients.

Methods that achieve a computational complexity that scales linearly with the number of degrees of freedom often rely on multiscale techniques, too. The *multi-grid method* [22, 9, 28] relaxes the error at different scales on coarser grids. *Algebraic multigrid methods* [33] try to achieve the robustness with respect to non-smooth coefficients by mostly heuristic strategies. Another successful class of methods, which are well-suited also for parallelization, are domain decomposition methods such as the *finite element tearing and interconnect method* [15, 14]. Although significant progress has been

---

\*This work was supported by the DFG collaborative research center SFB 1060.

<sup>†</sup>Institut für Numerische Simulation, Universität Bonn, Wegelerstrasse 6, 53115 Bonn, Germany

made in making the method robust with respect to variable coefficients (see [10, 27, 32, 30, 29]), the theory still contains assumptions on the coefficient's distribution.

The *mosaic skeleton method* [35] and *hierarchical ( $\mathcal{H}$ -) matrices* [23, 25] are historically related with the *fast multipole method* [19, 20], which can also be regarded as a multiscale method. Since  $\mathcal{H}$ -matrices are based on substructuring and low-rank approximation rather than multiscale effects, they are closer related with the *tree code algorithm* [1]. While fast multipole methods and the tree code algorithm are designed only for efficiently applying a non-local operator to a vector,  $\mathcal{H}$ -matrices provide an approximate algebra in which approximations of fully populated matrices (such as integral operators and the inverse or the factors of the LU decomposition of sparse matrices) can be computed with logarithmic-linear complexity. Although one could use  $\mathcal{H}$ -matrix approximations as direct solvers, it is usually more efficient to employ them as approximate preconditioners.

The existence of  $\mathcal{H}$ -matrix approximations to the inverse of finite element (FE) discretizations was proved in [6] for boundary value problems of scalar elliptic operators

$$\mathcal{L} := -\operatorname{div} A(x) \nabla \quad (1)$$

with a symmetric positive-definite coefficient  $A(x) \in \mathbb{R}^{d \times d}$  with  $a_{ij} \in L^\infty(\Omega)$ . The proof is based on the existence of exponentially convergent degenerate approximations

$$G_k(x, y) := \sum_{i=1}^k u_i(x) v_i(y) \approx G(x, y) \quad (2)$$

with suitable functions  $u_i, v_i$  to Green's function  $G$  for  $\mathcal{L}$  and the computational domain  $\Omega \subset \mathbb{R}^d$ . Due to the exponential convergence of the approximation (2), the accuracy  $\varepsilon$  enters the length  $k$  of the sum in (2) only logarithmically. Furthermore, the proof reveals that  $k$  does not depend on the smoothness of  $A$ . However, the bound on  $k$  still depends significantly on the contrast

$$\kappa_A := \frac{\lambda_{\max}}{\lambda_{\min}},$$

i.e. the ratio of the coefficient's largest and smallest eigenvalue  $\lambda_{\max} := \operatorname{ess\,sup}_{x \in \Omega} \lambda_{\max}(x)$  and  $\lambda_{\min} := \operatorname{ess\,inf}_{x \in \Omega} \lambda_{\max}(x)$ , where

$$\lambda_{\max}(x) := \max_{\lambda \in \sigma(A(x))} \lambda, \quad \lambda_{\min}(x) := \min_{\lambda \in \sigma(A(x))} \lambda,$$

and  $\sigma(A)$  denotes the spectrum of the matrix  $A$ . The size  $k$  determines the overall complexity of the  $\mathcal{H}$ -matrix approximation. Despite the dependence of  $k$  on  $\kappa_A$  in theory, an impact of high-contrast coefficients  $A$  on the efficiency of  $\mathcal{H}$ -matrices has never been observed in practise. The aim of this article is to underpin this observation theoretically. To this end, a norm will be introduced that depends on the coefficient  $A$  and generalizes the *flux norm* [8]. For the construction of approximations (2) it will be shown with respect to this norm that the Kolmogorov  $k$ -width of the space of  $\mathcal{L}$ -harmonic functions decays exponentially with  $k$  and does not depend on the contrast, i.e., it will be shown that

$$k \sim |\log \varepsilon|^{d+1},$$

whereas in [6] we proved  $k \sim \kappa_A^{d/2} |\log \varepsilon|^{d+1}$  with respect to the  $L^2$ -norm. While the focus of this article is on the dependence of  $k$  on the contrast in diffusion problems, the recent publication [13] analyzes  $k$  for high-frequency Helmholtz problems.

The approximation (2) of Green's function can be used to prove the existence of low-rank approximations to the inverse of FE discretizations. Since the inverse of  $\mathcal{L}$  has the representation

$$(\mathcal{L}^{-1}\varphi)(x) = \int_{\Omega} G(x, y)\varphi(y) \, dy, \quad x \in \Omega,$$

the existence of a degenerate approximation (2) on a pair of domains  $D_1 \times D_2 \subset \Omega \times \Omega$  leads to the existence of rank- $k$  approximations

$$(UV^T)_{ij} \approx (\mathcal{L}^{-1}\varphi_j, \varphi_i)_{L^2(\Omega)}, \quad U_{i\ell} := \int_{\Omega} u_{\ell}\varphi_i \, dx, \quad V_{j\ell} := \int_{\Omega} v_{\ell}\varphi_j \, dx,$$

to the FE discretization of  $\mathcal{L}^{-1}$  provided that the support of the FE basis functions  $\varphi_i$  and  $\varphi_j$  satisfy  $\text{supp } \varphi_i \subset D_1$  and  $\text{supp } \varphi_j \subset D_2$ . For the matrix approximation result, however, usual  $L^2$ -norm estimates are required as the matrix error is measured with respect to the spectral or the Frobenius norm. Changing the norm in the final estimate to the  $L^2$ -norm introduces a contrast-dependent term in the error estimate, which due to the exponential convergence enters the matrix rank  $k$  only logarithmically.

Note that the approximation technique presented in this article applies not only to operators (1). In [2] we considered general second order elliptic scalar operators and in [7] an analogous result was proved for the curl-curl operator. In practice, the LU factorization can be used to solve linear systems significantly faster than the inverse of a matrix. It is therefore important to remark that also the Schur complement and the factors of the LU decomposition can be approximated by hierarchical matrices with logarithmic-linear complexity; see [3] for a proof. The proof in [3] is based on the approximation of the inverse. Hence, the results on the approximation of the factors of the LU decomposition directly benefit from the new estimates of this article. Notice that hierarchical matrix approximations, from an algorithmic point of view, are constructed independently of the operator. Hence, this class of methods provides a fast and robust approach to problems with non-smooth, high-contrast coefficients.

The article is organized as follows. The way matrices are subdivided into sub-blocks is crucial for the efficiency of hierarchical matrices. Furthermore, the block structure is responsible for the properties of the domains on which Green's function is to be approximated. In Sect. 2, we will therefore shortly review the structure of hierarchical matrices. Sect. 3 contains the new low-dimensional approximation result with contrast-independent constant. To prove it, a suitable norm will be introduced in Sect. 3.1. The existence of degenerate approximations to Green's function on pairs of domains will be treated in Sect. 4. This will be based on interior regularity estimates and on the transfer property of flux norms. Numerical experiments in Sect. 5 support our theoretical findings.

## 2 Hierarchical matrices

The setting in which approximations of solution operators will be approximated in this article are hierarchical matrices. This methodology introduced by Hackbusch et al. [23, 25] is designed to handle

fully populated matrices such as approximations to the inverse or the factors of the LU decomposition with logarithmic-linear complexity; see [4, 24].

The efficiency of hierarchical matrices is based on low-rank approximations of each sub-matrix of a suitable partition  $P$  of the full set of matrix indices  $I \times I$ ,  $I := \{1, \dots, n\}$ . The construction of  $P$  has to account for two aims. On one hand, it has to guarantee that the rank  $k$  of the approximation

$$B_{ts} \approx XY^T, \quad X \in \mathbb{C}^{|t| \times k}, Y \in \mathbb{C}^{|s| \times k}, \quad (3)$$

to each block  $B_{ts}$ ,  $t \times s \in P$ , of a given matrix  $B \in \mathbb{C}^{I \times I}$  depends logarithmically on its approximation accuracy. Here,  $B$  denotes a fully populated matrix, e.g., the inverse of a stiffness matrix  $A$  resulting from FE discretization. On the other hand,  $P$  must be computable with logarithmic-linear complexity. The former issue will be addressed by the so-called *admissibility condition* in Sect. 2.2, while the latter problem can be solved by so-called *cluster trees*.

## 2.1 Cluster tree

Searching the set of possible partitions of  $I \times I$  for a partition  $P$  which guarantees (3) seems practically impossible since this set is considerably large. By restricting ourselves to blocks  $t \times s$  made up from rows  $t$  and columns  $s$  which are generated by recursive subdivision,  $P$  can be found with almost linear complexity. The structure which describes the way  $I$  is subdivided into smaller parts is the cluster tree. A tree  $T_I$  is called a *cluster tree* for an index set  $I$  if it satisfies the following conditions:

- (i)  $I$  is the root of  $T_I$ ;
- (ii) if  $t \in T_I$  is not a leaf, then  $t$  is a disjoint union of its sons  $S_I(t) = \{t_1, t_2\} \subset T_I$ .

We denote the set of leaves of the tree  $T_I$  by  $\mathcal{L}(T_I)$ .

A cluster tree for  $I$  can be computed, for instance, by the *bounding box method* or the *principal component analysis* [4]. The latter methods take into account the geometric information associated with the matrix indices. A nested dissection approach [5] based on the matrix graph often leads to significantly better results.

## 2.2 Block cluster tree

The approximation results from [6, 2] show that in order to be able to guarantee a sufficient approximation of each sub-matrix  $B_{ts}$ ,  $t \times s \in P$ , of  $B$  by a matrix of low rank, the sub-block  $t \times s$  has to satisfy the so-called *admissibility condition*

$$\min\{\text{diam } X_t, \text{diam } X_s\} \leq \eta \text{dist}(X_t, X_s) \quad (4)$$

for a given parameter  $\eta > 0$  or  $\min\{|t|, |s|\} \leq n_{\min}$  holds for a given block size parameter  $n_{\min} \in \mathbb{N}$ . Here,

$$\text{diam } X_t := \sup_{x, y \in X_t} |x - y| \quad \text{and} \quad \text{dist}(X_t, X_s) := \inf_{x \in X_t, y \in X_s} |x - y|$$

and the support  $X_t := \bigcup_{i \in t} X_i$  of a cluster  $t \in T_I$  is the union of the supports  $X_i := \text{supp } \varphi_i$  of the basis functions  $\varphi_i$ ,  $i \in t$ , corresponding to its indices. Notice that in order to satisfy (4), the distance

of the supports of  $t$  and  $s$  has to be large enough. This condition is caused by the fact that Green's functions of elliptic differential operators are singular for  $x = y$  only.

The partition is usually generated by recursive subdivision of  $I \times I$  descending the *block cluster tree*  $T_{I \times I}$ , which is a cluster tree for the set of matrix indices  $I \times I$  associated with the descendant mapping  $S_{I \times I}$  defined by

$$S_{I \times I}(t, s) := \begin{cases} \emptyset, & \text{if } S_I(t) = \emptyset \text{ or } S_I(s) = \emptyset, \\ S_I(t) \times S_I(s), & \text{else.} \end{cases}$$

The recursion stops in blocks which satisfy (4) or which are small enough. The set of leaves  $\mathcal{L}(T_{I \times I})$  of the block cluster tree  $T_{I \times I}$  forms a partition  $P$  of  $I \times I$ .

With a partition  $P$  constructed as above, the set of  $\mathcal{H}$ -matrices with blockwise rank  $k$  is defined as

$$\mathcal{H}(P, k) := \{M \in \mathbb{R}^{I \times I} : \text{rank } M_b \leq k \text{ for all } b \in P\}.$$

The storage requirement for  $B \in \mathcal{H}(P, k)$  is of the order  $kn \log n$ . Multiplying  $B$  by a vector can be done with  $\mathcal{O}(kn \log n)$  arithmetic operations. Since the sum of two  $\mathcal{H}$ -matrices  $B_1, B_2 \in \mathcal{H}(P, k)$  exceeds blockwise rank  $k$ , the sum has to be truncated to  $\mathcal{H}(P, k)$ . This can be done with complexity  $\mathcal{O}(k^2 n \log n)$  if an approximation error of controllable size can be tolerated. The complexity of computing an approximation to the product of two  $\mathcal{H}$ -matrices is  $\mathcal{O}(k^2 n (\log n)^2)$ ; see [23, 25, 18].

### 3 Weighted norms and interior regularity

The existence of finite-dimensional approximation spaces to the following space will turn out to be crucial for the existence of low-rank approximations of the discrete inverse.

Let  $D \subset \Omega$  be a domain. We investigate the approximation of functions from the space

$$X(D) = \{u \in H^1(D) : a(u, \varphi) = 0 \text{ for all } \varphi \in H_0^1(D), u = 0 \text{ on } \partial D \cap \partial \Omega\}$$

of  $\mathcal{L}$ -harmonic functions vanishing on  $\partial D \cap \partial \Omega$ , where  $a : H^1(D) \times H^1(D) \rightarrow \mathbb{R}$ ,

$$a(v, w) := \int_D \nabla w^T A \nabla v \, dx,$$

denotes the bilinear form associated with  $\mathcal{L}$ . The aim of this section is to construct a  $k$ -dimensional approximation space  $X_k \subset L^2(K)$  which provides  $A$ -independent error estimates, i.e., the Kolmogorov  $k$ -width of  $X(D)$  is bounded by

$$\sup_{u \in X(D)} \inf_{v \in X_k} \frac{\|u - v\|_K}{\|u\|_D} \leq \varepsilon_k \quad (5)$$

with some norm  $\|\cdot\|_K$  on  $K \subset D$  and  $\varepsilon_k \leq cq^{k/d}$ , where  $c > 0$  and  $0 < q < 1$  are independent of  $A$ . The construction of the space  $X_k$  will later be the basis for the construction of degenerate approximations (2).

In [6] we employed the Poincaré inequality on each piece  $K_i$ ,  $i = 1, \dots, k$ , of a sufficiently fine partition of  $K$ , i.e.  $\|u - \mu_i\|_{L^2(K_i)} \leq c_i \text{diam } K_i \|\nabla u\|_{L^2(K_i)}$  with some  $\mu_i \in \mathbb{R}$ ,  $i = 1, \dots, k$ , which leads to

$$\|u - \tilde{u}\|_{L^2(K)} \leq \varepsilon_k \|\nabla u\|_{L^2(K)}, \quad \varepsilon_k := \max_{i=1, \dots, k} c_i \text{diam } K_i,$$

with  $\tilde{u} \in X_k$  defined as  $\tilde{u}|_{K_i} := \mu_i$  and  $X_k$  the space of piecewise constant functions. The Caccioppoli-type inequality (see the remark after Lemma 1)

$$\|\nabla u\|_{L^2(K)} \leq \frac{c_T}{\sigma(K, D)} \|u\|_{L^2(D)} \quad (6)$$

leads to the desired estimate (5) provided that the distance

$$\sigma(K, D) := \text{dist}(K, \partial D \cap \Omega)$$

of  $K \subset D$  to  $\partial D$  within  $\Omega$  is positive.

The error estimate, however, is not  $A$ -independent as the  $A$ -dependence of  $c_T$  in (6) seems to be unavoidable. A remedy for the latter difficulty is to use weighted norms.

**Lemma 1.** *Let  $D \subset \Omega$ . Then for any set  $K \subset D$  satisfying  $\sigma(K, D) > 0$  it holds that*

$$\|A^{1/2} \nabla u\|_{L^2(K)} \leq \frac{2}{\sigma(K, D)} \|\lambda_{\max}^{1/2} u\|_{L^2(D)}$$

for all  $u \in X(D)$ .

*Proof.* Let  $\xi \in C^1(D)$  such that  $0 \leq \xi \leq 1$ ,  $\xi = 1$  in  $K$ ,  $\xi = 0$  in a neighborhood of  $\partial D \cap \Omega$  and

$$\|\nabla \xi\|_{\infty, D} \leq \frac{2}{\sigma(K, D)}.$$

We have  $\xi^2 u \in H^1(D)$  and  $\xi^2 u = 0$  on  $\partial D = (\partial D \cap \Omega) \cup (\partial D \cap \partial \Omega)$ . Hence,  $\xi^2 u$  can be used as a test function in the definition of  $X(D)$ , which leads to  $a(u, \xi^2 u) = 0$ . Using  $\nabla(\xi v) = \xi \nabla v + v \nabla \xi$  for  $v \in H^1(D)$ , from

$$\begin{aligned} (A \nabla u, \nabla(\xi^2 u))_{L^2(D)} &= (A \nabla u, \xi \nabla(\xi u) + \xi u \nabla \xi)_{L^2(D)} = (\xi A \nabla u, \nabla(\xi u) + u \nabla \xi)_{L^2(D)} \\ &= (A \nabla(\xi u) - u A \nabla \xi, \nabla(\xi u) + u \nabla \xi)_{L^2(D)} \\ &= (A^{1/2} \nabla(\xi u) - u A^{1/2} \nabla \xi, A^{1/2} \nabla(\xi u) + u A^{1/2} \nabla \xi)_{L^2(D)} \\ &= \|A^{1/2} \nabla(\xi u)\|_{L^2(D)}^2 - \|u A^{1/2} \nabla \xi\|_{L^2(D)}^2 \end{aligned}$$

we obtain that

$$\|A^{1/2} \nabla u\|_{L^2(K)}^2 \leq \|A^{1/2} \nabla(\xi u)\|_{L^2(D)}^2 = \|u A^{1/2} \nabla \xi\|_{L^2(D)}^2 \leq \|\lambda_{\max}^{1/2} u\|_{L^2(D)}^2 \|\nabla \xi\|_{\infty, D}^2.$$

□

**Remark.** Using  $\lambda_{\min}^{1/2} \|\nabla u\|_{L^2(K)} \leq \|A^{1/2} \nabla u\|_{L^2(K)}$ , Lemma 1 implies that

$$\|\nabla u\|_{L^2(K)} \leq \frac{2\sqrt{\kappa_A}}{\sigma(K, D)} \|u\|_{L^2(D)} \quad (7)$$

for all  $u \in X(D)$ .

Assume for a moment that  $A(x) = \alpha(x)I$  with some positive  $\alpha \in L^\infty(\Omega)$  and define

$$\|u\|_{L_\alpha^2(\Omega)} := \left( \int_\Omega \alpha |u|^2 dx \right)^{1/2}.$$

Due to Lemma 1, we obtain the following Caccioppoli inequality with  $\alpha$ -independent constant

$$\|\nabla u\|_{L_\alpha^2(K)} \leq \frac{2}{\sigma(K, D)} \|u\|_{L_\alpha^2(D)}, \quad u \in X(D).$$

Although the weight-independence of the Caccioppoli inequality can be achieved with respect to the  $\|\cdot\|_{L_\alpha^2}$ -norm, the independence of the Poincaré constant

$$\sup_{u \in H^1(K)} \inf_{\mu \in \mathbb{R}} \frac{\|u - \mu\|_{L_\alpha^2(K)}}{\|\nabla u\|_{L_\alpha^2(K)}}$$

of the weight  $\alpha$  is non-trivial. Weighted Poincaré inequalities with weight-independent constant have been presented in [16, 31] for the case that the domain  $K$  is partitioned into a finite number of disjoint Lipschitz domains  $K_i \subset K$ ,  $i = 1, \dots, r$ , and  $K_0 := K \setminus \bigcup_{i=1}^r \overline{K}_i$  in each of which the weight  $\alpha$  is constant, i.e.

$$\alpha(x) = \alpha_i, \quad x \in K_i,$$

with given numbers  $\alpha_i > 0$ ,  $i = 0, \dots, r$ . In [16], the case that  $K_i$  are inclusions of the domain  $K_0$  satisfying  $\alpha_i |K_i| \leq \alpha_0 |K_0|$ ,  $i = 1, \dots, r$ , is treated. It is proved that the Poincaré constant is weight-independent, but it depends on the number  $r$  of domains. In [31] a monotonicity of the sequence of coefficients is assumed in the spirit of [10, 27], for which a weight-independent Poincaré constant can be proved. The independence of  $\alpha$  does not hold in general. This can be seen from the following example.

**Example 1.** Let  $K = (-3, 3)^2 \subset \mathbb{R}^2$ ,  $K_1 = (-2, -1) \times (-1, 1)$ , and  $K_2 = (1, 2) \times (-1, 1)$ . For

$$\alpha(x) = \begin{cases} \delta^{-1}, & x \in K_1 \cup K_2, \\ 1, & \text{else,} \end{cases} \quad \text{and} \quad u(x) = \begin{cases} x_1, & |x_1| \leq 1, \\ 1, & 1 < x_1 < 3, \\ -1, & -3 < x_1 < -1, \end{cases}$$

with small  $\delta > 0$  we obtain for arbitrary  $\mu \in \mathbb{R}$

$$\int_K \alpha |u - \mu|^2 dx \geq \int_{K_1} \delta^{-1} (-1 - \mu)^2 dx + \int_{K_2} \delta^{-1} (1 - \mu)^2 dx = 4\delta^{-1} (1 + \mu^2) \geq 4\delta^{-1}$$

and

$$\int_K \alpha |\nabla u|^2 dx = 12,$$

which shows

$$\sup_{u \in H^1(K)} \inf_{\mu \in \mathbb{R}} \frac{\|u - \mu\|_{L_\alpha^2(K)}}{\|\nabla u\|_{L_\alpha^2(K)}} \geq \frac{1}{\sqrt{3\delta}}.$$

Therefore, weight-dependent Poincaré constants for general coefficients are unavoidable.

Our aim is to derive approximation estimates (5) which do not depend on the coefficient  $A$ . Since this cannot be achieved by a weighted Poincaré inequality for general coefficients, in the rest of this chapter a more sophisticated approach to the construction of a finite-dimensional approximation space  $X_k$  will be presented.

### 3.1 Flux norms

In this section we introduce a norm that will be useful for proving estimates of the form (5) with  $A$ -independent constants. Note that we consider arbitrary symmetric positive-definite coefficient matrices  $A(x) \in \mathbb{R}^{d \times d}$ . For  $v \in H^1(D)$  and any domain  $K \subset D$  with non-empty interior define the  $A$ -dependent function

$$\|v\|_{D,K} := \|\phi_v^D\|_{1,K},$$

where

$$\|\phi\|_{1,K}^2 := \|\phi\|_{L^2(K)}^2 + (\text{diam } K)^2 \|\nabla \phi\|_{L^2(K)}^2$$

and  $\phi_v^D \in H^1(D)$  denotes the solution of the Dirichlet problem

$$\begin{aligned} -\Delta \phi_v^D &= \mathcal{L}v \quad \text{in } D, \\ \phi_v^D &= v \quad \text{on } \partial D, \end{aligned}$$

i.e. we have  $\phi_v^D = v + z_v^D$ , where  $z_v^D \in H_0^1(D)$  satisfies

$$a_\Delta(z_v^D, \varphi) = a(v, \varphi) - a_\Delta(v, \varphi) \quad \text{for all } \varphi \in H_0^1(D), \quad (8)$$

where  $a_\Delta(u, v) := \int_\Omega \nabla u \nabla v dx$  denotes the bilinear form associated with the Laplacian. The existence and uniqueness of  $\phi_v^D \in H^1(D)$  follows from the Lax-Milgram theorem. In addition to  $\|\cdot\|_{D,K}$ , we introduce the  $A$ -dependent bilinear form  $(\cdot, \cdot)_{D,K} : H^1(D) \times H^1(D) \rightarrow \mathbb{R}$  as

$$(v, w)_{D,K} := (\phi_v^D, \phi_w^D)_{L^2(K)} + (\text{diam } K)^2 (\nabla \phi_v^D, \nabla \phi_w^D)_{L^2(K)},$$

which induces  $\|\cdot\|_{D,K}$ .

**Remark.** The function  $\|\cdot\|_{D,K}$  is a generalization of the so-called flux norm introduced in [8]. The latter is defined as  $\|\nabla \phi_v^D\|_{L^2(D)}$  for  $v \in H_0^1(D)$ . In the following,  $D$  will be fixed, while the dependence of  $\|\cdot\|_{D,K}$  on  $K$  is of particular importance for our needs. In particular, we cannot restrict ourselves to  $H_0^1(D)$ .



Notice that  $\|v\|_{D,K}$  evaluates  $\phi_v^D$  only on  $K \subset D$ . Hence,  $\|\cdot\|_{D,K}$  cannot be a norm on  $H^1(D)$ . The following lemma states that  $\|\cdot\|_{D,K}$  is a semi-norm on  $H^1(D)$  and a norm on  $X(D) \subset H^1(D)$ . This is due to the fact that  $v \in X(D)$  implies that  $\phi_v^D$  is harmonic and Lemma 8 (see Appendix) implies

$$\|v\|_{D,D} = \|\phi_v^D\|_{1,D} \leq c_{D,K} \|\phi_v^D\|_{1,K} = c_{D,K} \|v\|_{D,K}. \quad (9)$$

**Lemma 2.** *The bilinear form  $(\cdot, \cdot)_{D,K}$  is positive Hermitian (sometimes called semi-inner product) on  $H^1(D)$  and an inner product on  $X(D)$ .*

*Proof.* The symmetry of  $(\cdot, \cdot)_{D,K}$  is obvious. Its bilinearity follows from  $\phi_{\lambda v + w}^D = \lambda \phi_v^D + \phi_w^D$  for all  $\lambda \in \mathbb{R}$ ,  $v, w \in H^1(D)$ . For the positive definiteness assume that  $(v, v)_{D,K} = 0$  for  $v \in X(D)$ . From (9) we obtain that  $\phi_v^D = 0$  in  $D$ . Since  $\phi_v^D \in H^1(D)$ , also  $\phi_v^D|_{\partial D} = 0$ . Therefore,  $v \in H^1(D)$  satisfies  $\mathcal{L}v = 0$  in  $D$  and  $v = \phi_v^D = 0$  on  $\partial D$ , which yields  $v = 0$  in  $D$ .  $\square$

The semi-normed space  $(H^1(D), \|\cdot\|_{D,K})$  is a topological space induced by the semi-norm.

**Lemma 3.** *The space  $X(D)$  is closed in  $(H^1(D), \|\cdot\|_{D,K})$ .*

*Proof.* Let  $\{v_k\}_{k \in \mathbb{N}} \subset X(D)$  converge to  $v \in H^1(D)$  with respect to  $\|\cdot\|_{D,K}$ . From (9) we obtain

$$\|v_k - v\|_{D,D} \leq c \|v_k - v\|_{D,K} \rightarrow 0.$$

In particular, we have that  $\|\nabla(\phi_{v_k}^D - \phi_v^D)\|_{L^2(D)} \rightarrow 0$ . Hence, for  $\varphi \in H_0^1(D)$  it holds that

$$|a(v, \varphi)| = |a_\Delta(\phi_v^D, \varphi)| \leq \underbrace{|a_\Delta(\phi_{v_k}^D, \varphi)|}_{=0} + |a_\Delta(\phi_{v_k}^D - \phi_v^D, \varphi)| \leq \|\nabla(\phi_{v_k}^D - \phi_v^D)\|_{L^2(D)} \|\nabla \varphi\|_{L^2(D)} \rightarrow 0,$$

which shows that  $a(v, \varphi) = 0$ . Finally,  $v|_{\partial D \cap \partial \Omega} = \phi_v^D|_{\partial D \cap \partial \Omega} = 0$  proves  $v \in X(D)$ .  $\square$

The semi-inner product  $(\cdot, \cdot)_{D,K}$  on  $H^1(D)$  is sufficient to define an element of best approximation in the closed subspace  $X(D)$  (cf. Lemma 3) of  $H^1(D)$ . The positive definiteness of  $(\cdot, \cdot)_{D,K}$  on  $X(D)$  (cf. Lemma 2) implies that any element of best approximation is unique. Hence, the  $(\cdot, \cdot)_{D,K}$ -orthogonal projection

$$\mathcal{P}_{D,K} : H^1(D) \rightarrow X(D)$$

is well-defined.

The following equivalence relation will be useful at the end of this section, when error estimates with respect to  $\|\cdot\|_{D,D}$  are reformulated in the usual  $L^2(D)$ -norm.

**Lemma 4.** *There are constants  $c_1, c_2 > 0$  (depending on  $A$ ) such that*

$$\frac{1}{c_1} \|v\|_{1,D} \leq \|v\|_{D,D} \leq c_2 \|v\|_{1,D}$$

for all  $v \in H^1(D)$ . It holds that  $c_1 \sim \kappa_A$  and  $c_2 \sim \lambda_{\max}$ .

*Proof.* The choice  $\varphi = z_v^D$  in (8) shows

$$\|\nabla z_v^D\|_{L^2(D)}^2 = \int_D \nabla v^T (A - I) \nabla z_v^D \, dx \leq |\lambda_{\max} - 1| \|\nabla v\|_{L^2(D)} \|\nabla z_v^D\|_{L^2(D)}$$

and hence  $\|\nabla z_v^D\|_{L^2(D)} \leq |\lambda_{\max} - 1| \|\nabla v\|_{L^2(D)}$ . With  $c_A := |\lambda_{\max} - 1| + 1 \leq \max\{2, \lambda_{\max}\}$  we obtain

$$\|\nabla \phi_v^D\|_{L^2(D)} \leq \|\nabla z_v^D\|_{L^2(D)} + \|\nabla v\|_{L^2(D)} \leq c_A \|\nabla v\|_{L^2(D)}. \quad (10)$$

In addition, Poincaré's inequality leads to

$$\|z_v^D\|_{L^2(D)} \leq c_P \operatorname{diam} D \|\nabla z_v^D\|_{L^2(D)} \leq c_P \operatorname{diam} D |\lambda_{\max} - 1| \|\nabla v\|_{L^2(D)},$$

which shows

$$\|\phi_v^D\|_{L^2(D)} \leq \|v\|_{L^2(D)} + c_P c_A \operatorname{diam} D \|\nabla v\|_{L^2(D)}. \quad (11)$$

From (10) and (11) we obtain

$$\begin{aligned} \|v\|_{D,D}^2 &\leq (\|v\|_{L^2(D)} + c_P c_A \operatorname{diam} D \|\nabla v\|_{L^2(D)})^2 + c_A^2 (\operatorname{diam} D)^2 \|\nabla v\|_{L^2(D)}^2 \\ &\leq 2\|v\|_{L^2(D)}^2 + 2c_P^2 c_A^2 (\operatorname{diam} D)^2 \|\nabla v\|_{L^2(D)}^2 + c_A^2 (\operatorname{diam} D)^2 \|\nabla v\|_{L^2(D)}^2 \leq c_2^2 \|v\|_{1,D}^2 \end{aligned}$$

if we set  $c_2 := \max\{\sqrt{2}, c_A \sqrt{2c_P^2 + 1}\}$ . Similarly, testing (8) with  $\varphi = z_v^D$  one has  $a(z_v^D, z_v^D) = a(\phi_v^D, z_v^D) - a_\Delta(\phi_v^D, z_v^D)$  and thus

$$\begin{aligned} \|v\|_{1,D}^2 &\leq \left( \|\phi_v^D\|_{L^2(D)} + \frac{c_P c_A}{\lambda_{\min}} \operatorname{diam} D \|\nabla \phi_v^D\|_{L^2(D)} \right)^2 + \left( \frac{c_A}{\lambda_{\min}} \right)^2 (\operatorname{diam} D)^2 \|\nabla \phi_v^D\|_{L^2(D)}^2 \\ &\leq 2\|\phi_v^D\|_{L^2(D)}^2 + \left( \frac{c_A}{\lambda_{\min}} \right)^2 (2c_P^2 + 1) (\operatorname{diam} D)^2 \|\nabla \phi_v^D\|_{L^2(D)}^2 \leq c_1^2 \|v\|_{D,D}^2 \end{aligned}$$

with  $c_1 := \max\{\sqrt{2}, \frac{c_A}{\lambda_{\min}} \sqrt{2c_P^2 + 1}\}$ . □

### 3.2 Approximation from finite-dimensional spaces

The approximation of a given element  $u \in X(D)$  with respect to  $\|\cdot\|_{D,K}$  can be related to the approximation of the harmonic function  $\phi_u^D$  with respect to  $\|\cdot\|_{1,K}$ .

**Lemma 5.** *Let  $V_p = \operatorname{span}\{\varphi_1, \dots, \varphi_p\}$ ,  $W_p = \operatorname{span}\{\psi_1, \dots, \psi_p\}$  be  $p$ -dimensional subspaces of  $H^1(D)$  such that*

$$\mathcal{L}\varphi_i = -\Delta\psi_i \quad \text{and} \quad \varphi_i|_{\partial D} = \psi_i|_{\partial D}, \quad i = 1, \dots, p. \quad (12)$$

*Then*

$$\inf_{v \in V_p} \|u - \mathcal{P}_{D,K} v\|_{D,K} \leq \inf_{w \in W_p} \|\phi_u^D - w\|_{1,K}$$

*for all  $K \subset D$ .*

*Proof.* Let  $w = \sum_{i=1}^p \gamma_i \psi_i \in W_p$  with  $\gamma_i \in \mathbb{R}$ . Then  $v := \sum_{i=1}^p \gamma_i \varphi_i \in V_p$  satisfies  $\mathcal{L}v = -\Delta w$  in  $D$  and  $v = w$  on  $\partial D$ . Hence,  $\phi_v^D = w$  and from  $\phi_{u-v}^D = \phi_u^D - \phi_v^D$  we obtain

$$\|u - \mathcal{P}_{D,K}v\|_{D,K} = \|\mathcal{P}_{D,K}(u - v)\|_{D,K} \leq \|u - v\|_{D,K} = \|\phi_u^D - \phi_v^D\|_{1,K} = \|\phi_u^D - w\|_{1,K},$$

which proves that  $\inf_{v \in V_p} \|u - \mathcal{P}_{D,K}v\|_{D,K} \leq \|\phi_u^D - w\|_{1,K}$ . The assertion follows since  $w \in W_p$  is arbitrary.  $\square$

The latter property will now be used to construct finite-dimensional spaces  $X_k$  which approximate  $X(D)$  independently of  $A$ . For deriving error estimates it is required that  $\phi_u^D$  has a slightly higher regularity. The following interior regularity result is proved similar to [17, Thm. 8.8]. While the latter result expects a positive distance  $\text{dist}(K, \partial D)$ , our modification assumes only  $\sigma(K, D) > 0$ .

**Lemma 6.** *For any subset  $K \subset D$  satisfying  $\sigma := \sigma(K, D) > 0$  it holds that  $\phi_u^D \in H^2(K)$  and*

$$\|\partial_i \nabla \phi_u^D\|_{L^2(K)} \leq \frac{2}{\sigma} \|\partial_i \phi_u^D\|_{L^2(D)}, \quad i = 1, \dots, d.$$

*Proof.* Due to  $u \in X(D)$ , we have that  $\phi_u^D \in H^1(D)$  is harmonic and  $\phi_u^D = 0$  on  $\partial D \cap \partial \Omega$ . Let  $\xi \in C^\infty(D)$  satisfy  $0 \leq \xi \leq 1$ ,  $\xi = 1$  in  $K$ ,  $\xi = 0$  in a neighborhood of  $\partial D \cap \Omega$  and

$$\|\nabla \xi\|_{\infty, D} \leq \frac{2}{\sigma}.$$

Since  $\xi^2 \phi_u^D = 0$  on  $\partial D = (\partial D \cap \Omega) \cup (\partial D \cap \partial \Omega)$ , we have  $\hat{K} := \text{supp } \xi^2 \phi_u^D \subset D$ . For  $0 < h < \frac{1}{2} \text{dist}(\hat{K}, \partial D)$  define  $\varphi := \partial_i^{-h}(\xi^2 \partial_i^h \phi_u^D) \in H_0^1(D)$ , where

$$\partial_i^h u(x) := \frac{u(x + h e_i) - u(x)}{h}$$

denotes the difference quotient of  $u$  in direction  $i$ . Due to the dense imbedding of  $C_0^\infty(D)$  in  $H_0^1(D)$ , we may assume that  $\varphi \in C_0^\infty(D)$ . Since  $\phi_u^D$  is harmonic, from

$$\begin{aligned} 0 &= - \int_D \nabla \phi_u^D \cdot \nabla [\partial_i^{-h}(\xi^2 \partial_i^h \phi_u^D)] \, dx = \int_D \nabla \partial_i^h \phi_u^D \cdot \nabla [\xi^2 \partial_i^h \phi_u^D] \, dx \\ &= \int_D \nabla \partial_i^h \phi_u^D \cdot [\xi \nabla(\xi \partial_i^h \phi_u^D) + \xi \nabla \xi \partial_i^h \phi_u^D] \, dx \\ &= \int_D [\nabla(\xi \partial_i^h \phi_u^D) - \nabla \xi \partial_i^h \phi_u^D] \cdot [\nabla(\xi \partial_i^h \phi_u^D) + \nabla \xi \partial_i^h \phi_u^D] \, dx \\ &= \int_D |\nabla(\xi \partial_i^h \phi_u^D)|^2 - |\nabla \xi \partial_i^h \phi_u^D|^2 \, dx \end{aligned}$$

we obtain that

$$\|\nabla \partial_i^h \phi_u^D\|_{L^2(K)}^2 \leq \int_D |\nabla(\xi \partial_i^h \phi_u^D)|^2 \, dx = \int_D |\nabla \xi \partial_i^h \phi_u^D|^2 \, dx \leq \frac{4}{\sigma^2} \|\partial_i^h \phi_u^D\|_{L^2(\hat{K})}^2.$$

Hence, we have shown the desired estimate for the finite differences. The estimate for the differential operators follow from applying two results from [17]. With Lemma 7.23 from [17] we obtain  $\|\partial_i^h \phi_u^D\|_{L^2(\hat{K})} \leq \|\partial_i \phi_u^D\|_{L^2(D)}$ . By Lemma 7.24 from [17] it follows that  $\phi_u^D \in H^2(K)$  and

$$\|\partial_i \nabla \phi_u^D\|_{L^2(K)} \leq \frac{2}{\sigma} \|\partial_i \phi_u^D\|_{L^2(D)}.$$

□

Assume that  $\Omega$  is polyhedral, and let  $\Delta_H$  be a quasi-uniform polyhedrization of  $\Omega$  with mesh size  $H > 0$ . We define the space

$$W_p := \{v|_D, v \in \mathcal{S}^{1,0}(\Delta_H)\}$$

of piecewise linear finite elements  $\mathcal{S}^{1,0}(\Delta_H)$  restricted to  $D$ . Obviously, the mesh size  $H$  is connected with the dimension  $p$  of  $W_p$  as

$$H \leq c_R \frac{\text{diam } D}{\sqrt[p]{p}} \quad (13)$$

with a constant  $c_R > 0$ . Let  $K_H \subset \Delta_H$  be the smallest polyhedrization such that  $K \subset K_H \subset D$  and  $\sigma(K_H, D) > 0$ . Note that we are interested in the limit  $H \rightarrow 0$ . Due to  $\phi_u^D \in H^2(K_H)$  (see Lemma 6), the nodal interpolation operator  $\mathfrak{I}_H : L^2(D) \rightarrow W_p$  provides the following interpolation error estimates

$$\|\phi_u^D - \mathfrak{I}_H \phi_u^D\|_{L^2(K_H)} \leq c_{\mathfrak{I}} H |\phi_u^D|_{H^1(K_H)}, \quad \|\phi_u^D - \mathfrak{I}_H \phi_u^D\|_{H^1(K_H)} \leq c_{\mathfrak{I}} H |\phi_u^D|_{H^2(K_H)}. \quad (14)$$

Let another set  $K' \subset D$  satisfy  $K \subset K_H \subset K'$  such that  $\sigma(K, K') > 0$ . We set

$$\rho := \frac{\text{diam } D}{\sigma(K, K')}.$$

Using  $\sigma(K_H, K') \geq \sigma(K, K') - H$  and  $H \leq c_R p^{-1/d} \text{diam } D = c_R p^{-1/d} \rho \sigma(K, K')$ , we obtain

$$\sigma := \sigma(K_H, K') \geq \left(1 - \frac{c_R \rho}{\sqrt[p]{p}}\right) \sigma(K, K') \geq \frac{1}{2} \sigma(K, K')$$

for  $p \geq (2c_R \rho)^d$ . From (14), Lemma 1 applied to  $A = I$ , and Lemma 6 we obtain

$$\begin{aligned} \inf_{w \in W_p} \|\phi_u^D - w\|_{1, K_H}^2 &\leq c_{\mathfrak{I}}^2 H^2 \left( |\phi_u^D|_{H^1(K_H)}^2 + (\text{diam } K_H)^2 |\phi_u^D|_{H^2(K_H)}^2 \right) \\ &\leq 4 \frac{c_{\mathfrak{I}}^2 H^2}{\sigma^2} \left( \|\phi_u^D\|_{L^2(K')}^2 + d (\text{diam } K_H)^2 \|\nabla \phi_u^D\|_{L^2(K')}^2 \right) \end{aligned}$$

and hence

$$\inf_{w \in W_p} \|\phi_u^D - w\|_{1, K} \leq \inf_{w \in W_p} \|\phi_u^D - w\|_{1, K_H} \leq \frac{4dc_{\mathfrak{I}}H}{\sigma(K, K')} \|\phi_u^D\|_{1, K'}. \quad (15)$$

If  $\psi_1, \dots, \psi_p$  denotes a basis of  $W_p$  and  $\varphi_1, \dots, \varphi_p$  are the corresponding solutions of (12), then Lemma 5, (13), and (15) show the existence of an at most  $p$ -dimensional space

$$Y_K := \mathcal{P}_{D, K} V_p \subset X(D)$$

with  $p \geq (2c_R\rho)^d$  such that

$$\inf_{v \in Y_K} \|u - v\|_{D,K} \leq \frac{4dc_3H}{\sigma(K, K')} \|\phi_u^D\|_{1,K'} \leq \frac{4dc_3c_R}{\sqrt[d]{p}} \rho \|\phi_u^D\|_{1,K'} = \frac{c_S}{\sqrt[d]{p}} \rho \|u\|_{D,K'} \quad (16)$$

with the  $A$ -independent constant  $c_S := 4dc_3c_R$ .

For the following theorem the algebraic decay (16) is exploited recursively on a sequence of nested domains to obtain the desired exponential convergence.

**Theorem 1.** *Let  $K \subset D$  such that  $\eta\sigma(K, D) \geq \text{diam } D$  with some  $\eta > 0$ . For any  $\varepsilon > 0$  there is a  $k$ -dimensional subspace  $X_k \subset X(D)$  satisfying*

$$\inf_{v \in X_k} \|u - v\|_{D,K} \leq \varepsilon \|u\|_{D,D} \quad \text{for all } u \in X(D) \quad (17)$$

provided that  $k \geq c_\eta \lceil \log \varepsilon \rceil^{d+1}$ , where  $c_\eta := \lceil \eta \max\{c_{Se}, 2c_R\} \rceil^d$ .

*Proof.* Let  $\ell = \lceil \log \varepsilon \rceil$  and  $r_0 := \sigma(K, D)$ . We consider a sequence of nested domains

$$K_j = \{x \in \Omega : \text{dist}(x, K) \leq r_0(\ell - j)/\ell\}, \quad j = 0, \dots, \ell.$$

Notice that  $K = K_\ell \subset K_{\ell-1} \subset \dots \subset K_0 \subset D$  with

$$\sigma(K_j, K_{j-1}) = \frac{r_0}{\ell} = \frac{\sigma(K, D)}{\ell} \geq \frac{\text{diam } D}{\eta\ell}, \quad j = 1, \dots, \ell.$$

According to (16) (with the choice  $K := K_j$  and  $K' := K_{j-1}$ ), there is  $Y_{K_j} \subset X(D)$ ,  $\dim Y_{K_j} \leq p$ , so that for all  $u \in X(D)$

$$\inf_{v \in Y_{K_j}} \|u - v\|_{D,K_j} \leq \frac{c_S}{\sqrt[d]{p}} \eta\ell \|u\|_{D,K_{j-1}} \leq \varepsilon^{1/\ell} \|u\|_{D,K_{j-1}}, \quad (18)$$

if we choose  $p \geq p_0 := \lceil \eta\ell \max\{c_S\varepsilon^{-1/\ell}, 2c_R\} \rceil^d$ .

Let  $e_0 := u \in X(D)$ . Estimate (18) defines an element  $v_1 \in Y_{K_1}$  and hence  $e_1 := e_0 - v_1 \in X(D)$  so that

$$\|e_1\|_{D,K_1} \leq \varepsilon^{1/\ell} \|e_0\|_{D,K_0}.$$

Similarly, from (18) we obtain approximants  $v_j \in Y_{K_j}$ ,  $j = 2, \dots, \ell$ , so that with  $e_j := e_{j-1} - v_j$

$$\|e_j\|_{D,K_j} \leq \varepsilon^{1/\ell} \|e_{j-1}\|_{D,K_{j-1}}.$$

Since  $e_0 = e_\ell + \sum_{j=1}^\ell v_j$  and  $\sum_{j=1}^\ell v_j \in X_k := \bigoplus_{j=1}^\ell Y_{K_j}$ , we are led to

$$\inf_{v \in X_k} \|e_0 - v\|_{D,K_\ell} \leq \|e_\ell\|_{D,K_\ell} \leq (\varepsilon^{1/\ell})^\ell \|e_0\|_{D,K_0} \leq \varepsilon \|u\|_{D,D}.$$

The dimension  $k$  of  $X_k$  is bounded by  $p\ell$ . From  $\varepsilon^{-1/\ell} \leq e$  we obtain that

$$p_0 \leq \lceil \eta\ell \max\{c_{Se}, 2c_R\} \rceil^d,$$

which proves the assertion. □

## 4 Separable approximation of solution operators

In [6, 2] we were able to prove that inverse FE stiffness matrices of scalar elliptic boundary value problems can be approximated using hierarchical matrices with logarithmic-linear complexity. Since we do not want to repeat the proofs from [6], we concentrate on the central problem of constructing degenerate kernel expansions

$$G_k(x, y) := \sum_{i=1}^k u_i(x) v_i(y)$$

to Green's function  $G$  for the operator  $\mathcal{L}$  satisfying

$$(i) \quad \mathcal{L}G(x, \cdot) = \delta_x \text{ in } \Omega,$$

$$(ii) \quad \mathcal{L}G(x, \cdot) = 0 \text{ on } \partial\Omega$$

for all  $x \in \Omega$ ; for the existence of  $G$  see [21]. The rest of the proof in [6] is based on the following existence result (see Theorem 3.5 in [2]) and can be applied without changes.

**Theorem 2.** *Let  $D_1, D_2 \subset \Omega$  be two domains such that  $D_2$  is convex and*

$$\eta \operatorname{dist}(D_1, D_2) \geq \operatorname{diam} D_2. \quad (19)$$

*Then for any  $\varepsilon > 0$  there is a separable approximation*

$$G_k(x, y) = \sum_{i=1}^k u_i(x) v_i(y) \quad \text{with } k \leq c_{\eta, A} |\log \varepsilon|^{d+1}$$

*satisfying*

$$\|G(x, \cdot) - G_k(x, \cdot)\|_{L^2(D_2)} \leq \varepsilon \|G(x, \cdot)\|_{L^2(\hat{D}_2)} \quad \text{for all } x \in D_1,$$

*where  $\hat{D}_2 := \{y \in \Omega : 2\eta \operatorname{dist}(y, D_2) \leq \operatorname{diam} D_2\}$ .*

It is remarkable that this result holds for arbitrary coefficients  $a_{ij} \in L^\infty(\Omega)$  of the operator  $\mathcal{L}$  satisfying  $\lambda_{\min} > 0$  because in this case Green's function does not possess any higher regularity. Although the approximation is independent of the smoothness of the coefficient  $A$ , it can be seen from the proof in [2] that the estimate on  $k$  (i.e. the constant  $c_{\eta, A}$ ) depends on the contrast  $\kappa_A$  as  $\kappa_A^{d/2}$ .

In this section it will be proved using the  $\|\cdot\|_{D, K}$ -norm that the dependence on  $\kappa_A$  can be avoided. Note that the approximation technique presented in this article applies not only to operators (1). In [2] we considered general second order elliptic scalar operators and in [7] an analogous result was proved for the curl-curl operator. For practical reasons it is important to remark that also the Schur complement and the factors of the LU decomposition can be handled by hierarchical matrices with logarithmic-linear complexity. This paves the way to approximate direct methods that do not suffer from the well-known effect of fill-in. The proof [3] is based on the approximation of the inverse. Hence, also the results in [3] directly benefit from the new estimates of this article.

The next theorem is a variant of Theorem 2, in which the approximation error (or the rank of the approximation) does not depend on the coefficient  $A$ . To this end, the results obtained in the previous sections for arbitrary elements in  $X(D)$  will now be applied to a particular element. Although the entries of  $A$  are only measurable, it can be proved that  $g_x := G(x, \cdot) \in H^1(\Omega \setminus B_\varepsilon(x))$  for all  $\varepsilon > 0$ ; see [21]. Hence,  $g_x|_D$  is in  $X(D)$  for all  $x \in \Omega \setminus \bar{D}$ .

**Theorem 3.** Let  $D_1, D_2 \subset \Omega$  satisfy (19). Then for any  $\varepsilon > 0$  there is a separable approximation

$$G_k(x, y) = \sum_{i=1}^k u_i(x) v_i(y) \quad \text{with } k \leq c_\eta \lceil |\log \varepsilon| \rceil^{d+1},$$

so that for all  $x \in D_1$

$$\|G(x, \cdot) - G_k(x, \cdot)\|_{\hat{D}_2, D_2} \leq \varepsilon \|G(x, \cdot)\|_{\hat{D}_2, \hat{D}_2}, \quad (20)$$

where  $\hat{D}_2$  is defined in Theorem 2 and  $c_\eta := \lceil 2(\eta + 1) \max\{c_{Se}, 2c_R\} \rceil^d$ .

*Proof.* Notice that because of  $\text{dist}(D_1, \hat{D}_2) > 0$ , we have  $g_x|_{\hat{D}_2} \in X(\hat{D}_2)$  for all  $x \in D_1$ . Since  $\text{diam } \hat{D}_2 \leq (1 + 1/\eta) \text{diam } D_2$ , it holds that

$$\sigma(D_2, \hat{D}_2) = \frac{1}{2\eta} \text{diam } D_2 \geq \frac{1}{2(\eta + 1)} \text{diam } \hat{D}_2.$$

Hence, Theorem 1 can be applied with  $K := D_2$ ,  $D := \hat{D}_2$ , and  $\eta' = 2(\eta + 1)$ .

Let  $\{v_1, \dots, v_k\}$  be a basis of the subspace  $X_k \subset X(\hat{D}_2)$  with  $k \geq c_{\eta'} \lceil |\log \varepsilon| \rceil^{d+1}$ . By means of (17) we can decompose  $g_x$  as  $g_x = \tilde{g}_x + e_x$  with  $\tilde{g}_x \in X_k$  and

$$\|e_x\|_{\hat{D}_2, D_2} \leq \varepsilon \|g_x\|_{\hat{D}_2, \hat{D}_2}.$$

Expressing  $\tilde{g}_x$  in the basis of  $X_k$ , we obtain

$$\tilde{g}_x = \sum_{i=1}^k u_i(x) v_i$$

with coefficients  $u_i(x)$  depending on  $x \in D_1$ . The function  $G_k(x, y) := \tilde{g}_x(y)$  satisfies (20).  $\square$

The previous theorem shows that  $k$  is bounded independently of the coefficient  $A$ . Estimate (20), however, is an estimate with respect to the  $\|\cdot\|_{D, K}$ -norm. If the arguments from [6] are to be applied in order to estimate the rank of  $\mathcal{H}$ -matrix approximations, the usual  $L^2$ -norm (as in Theorem 2) has to be used.

**Theorem 4.** Let  $D_1, D_2 \subset \Omega$  satisfy (19). Then for any  $\varepsilon > 0$  there is a separable approximation

$$G_k(x, y) = \sum_{i=1}^k u_i(x) v_i(y) \quad \text{with } k \leq c_\eta \lceil |\log(\varepsilon \kappa_A^{-3/2} / \lambda_{\max})| \rceil^{d+1},$$

so that for all  $x \in D_1$

$$\|G(x, \cdot) - G_k(x, \cdot)\|_{L^2(D_2)} \leq \varepsilon \|G(x, \cdot)\|_{L^2(\hat{D}_2)},$$

where  $\hat{D}_2$  is defined in Theorem 2 and  $c_\eta := \lceil (4\eta + 2) \max\{c_{Se}, 2c_R\} \rceil^d$ .

*Proof.* As in the proof of Theorem 3 we have  $g_x|_{\hat{D}_2} \in X(\hat{D}_2)$  for all  $x \in D_1$ . Let

$$D'_2 := \{y \in \Omega : 4\eta \operatorname{dist}(y, D_2) \leq \operatorname{diam} D_2\}.$$

Then  $D_2 \subset D'_2 \subset \hat{D}_2$ ,  $\operatorname{diam} D'_2 \leq (1 + 1/(2\eta)) \operatorname{diam} D_2$ , and

$$\sigma(D_2, D'_2) = \frac{1}{4\eta} \operatorname{diam} D_2 \geq \frac{1}{4\eta + 2} \operatorname{diam} D'_2.$$

Hence, Theorem 1 can be applied with  $K := D_2$ ,  $D := D'_2$ , and  $\eta' = 4\eta + 2$ , which yields the subspace  $X_k \subset X(D'_2)$  with  $k \geq c_{\eta'} \lceil |\log \varepsilon| \rceil^{d+1}$ . By means of (17) we can decompose  $g_x$  as  $g_x = \tilde{g}_x + e_x$  with  $\tilde{g}_x \in X_k$  and  $e_x \in X(D'_2)$  such that

$$\begin{aligned} \|e_x\|_{L^2(D_2)} &\leq c_1 \|e_x\|_{D'_2, D_2} \leq c_1 c_{D'_2, D_2} \|e_x\|_{D'_2, D_2} \\ &\leq c_1 c_{D'_2, D_2} \varepsilon \|g_x\|_{D'_2, D'_2} \leq c_1 c_2 c_{D'_2, D_2} \varepsilon \|g_x\|_{1, D'_2} \\ &= c_1 c_2 c_{D'_2, D_2} \varepsilon \sqrt{\|g_x\|_{L^2(D'_2)}^2 + (\operatorname{diam} D'_2)^2 \|\nabla g_x\|_{L^2(D'_2)}^2}, \end{aligned}$$

where we used Lemma 4 and (9). From (7) we obtain that

$$\|\nabla g_x\|_{L^2(D'_2)} \leq \frac{2\sqrt{\kappa_A}}{\sigma(D'_2, \hat{D}_2)} \|g_x\|_{L^2(\hat{D}_2)}.$$

The dependence of  $c_{D,K}$  on the domains  $D$  and  $K$  is not explicitly known. A scaling argument reveals that

$$c_{D,K} \leq \hat{c}_{D,K} \left( \frac{\operatorname{diam} D}{\operatorname{diam} K} \right)^{d/2},$$

where  $\hat{c}_{D,K} > 0$  depends on the shapes of  $D$  and  $K$  but not on their diameters. Due to  $\sigma(D'_2, \hat{D}_2) = \sigma(D_2, D'_2)$ , we have

$$\begin{aligned} \frac{\operatorname{diam} D'_2}{\sigma(D'_2, \hat{D}_2)} &= 4\eta + 2, \\ c_{D'_2, D_2} &\leq \hat{c}_{D'_2, D_2} \left( \frac{\operatorname{diam} D'_2}{\operatorname{diam} D_2} \right)^{d/2} \leq \hat{c}_{D'_2, D_2} \left( 1 + \frac{1}{2\eta} \right)^{d/2} \end{aligned}$$

and we obtain from  $c_1 \sim \kappa_A$  and  $c_2 \sim \lambda_{\max}$  that

$$\begin{aligned} \|e_x\|_{L^2(D_2)} &\leq c_1 c_2 \hat{c}_{D'_2, D_2} \left( 1 + \frac{1}{2\eta} \right)^{d/2} \varepsilon \sqrt{1 + 4\kappa_A(4\eta + 2)^2} \|g_x\|_{L^2(\hat{D}_2)} \\ &\leq c \lambda_{\max} \kappa_A^{3/2} \varepsilon \|g_x\|_{L^2(\hat{D}_2)}. \end{aligned}$$

The assertion follows with the same arguments as used in the proof of Theorem 3.  $\square$

Compared with the “old” Theorem 2, in which  $k$  depends on  $\kappa_A$  as  $k \sim \kappa_A^{d/2} |\log \varepsilon|^{d+1}$ , the new proof via  $A$ -dependent norms  $\|\cdot\|_{D,K}$  significantly improves the rank estimate in the  $L^2$ -norm to

$$k \sim |\log(\varepsilon \kappa_A^{-3/2} / \lambda_{\max})|^{d+1},$$

i.e., the contrast  $\kappa_A$  enters the complexity of  $\mathcal{H}$ -matrices only logarithmically.



## 5 Numerical experiments

In this section the influence of the coefficient  $A$  in the differential operator (1) will be investigated numerically on the unit square  $\Omega := [0, 1]^3$  in  $\mathbb{R}^3$ . To this end, we choose  $r$  balls  $\Omega_i = B_{s_i}(x_i)$ ,  $i = 1, \dots, r$ , centered at randomly generated centers  $x_i \in \Omega$ ; see Fig. 1. For the coefficient  $A = \alpha I$

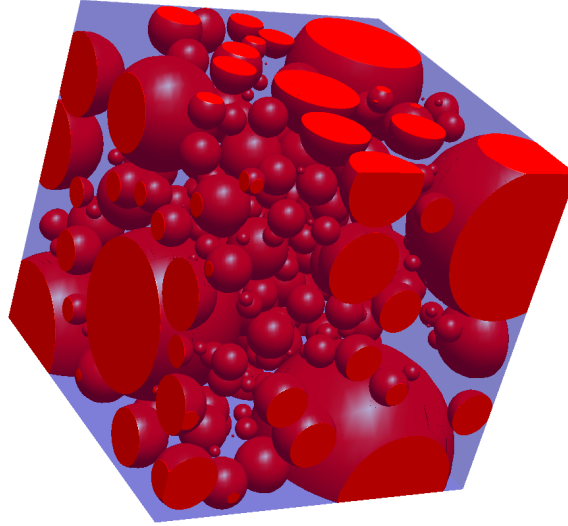


Figure 1:  $p = 200$  inclusions with random coefficients in  $[0, M]$ .

we use

$$\alpha(x) = \begin{cases} \alpha_i, & x \in \Omega_i, \\ 1, & \text{else,} \end{cases}$$

where each  $\alpha_i \in [0, M]$  is randomly chosen. Table 1 contains the storage requirement per degree of freedom when approximating the inverse of standard FE discretizations using hierarchical matrices with relative accuracy  $\varepsilon = 10^{-3}$ . The emphasis of these tests is not on the scaling behavior, i.e. the dependence of the complexity with respect to the number of degrees of freedom  $n$ . Such tests were published in [6, 2, 7] and in many other articles. Here, we are primarily interested in the dependence of the rank  $k$  on the number of domains  $r$  and on the amplitude  $M$  of the coefficient  $\alpha$ . Therefore, only one choice of  $n$  will be considered. The table shows the storage requirement of the hierarchical matrix approximation to the inverse. Since the storage depends linearly on  $k$ , this allows a direct comparison. Apparently, the storage requirement is bounded with respect to the number of domains  $r$

$p \setminus M$	1	10	100	1000	10000
1	22.92	22.84	22.63	22.54	22.07
10	22.54	22.69	22.42	18.94	17.34
100	22.74	22.84	23.58	18.15	15.51
1000	22.83	22.59	22.12	16.00	10.78

Table 1: Storage per degree of freedom in kByte for  $n = 166\,375$ .

and with respect to the amplitude  $M$ . The complexity actually improves with  $p$  and  $M$ .

## Appendix

Let  $D \subset \mathbb{R}^d$  be a bounded domain and let  $K \subset D$  be a subdomain. We denote the dual space of  $H^1(D)$  by  $\tilde{H}^{-1}(D)$ .

**Lemma 7.** *The Newton potential operator  $N_K : \tilde{H}^{-1}(D) \rightarrow H^1(D)$  defined by*

$$(N_K \varphi)(y) := \int_K S(x, y) \varphi(x) \, dx, \quad y \in D,$$

with

$$S(x, y) := \frac{1}{4\pi} \frac{1}{|x - y|}$$

is continuous, i.e.  $\|N_K \varphi\|_{H^1(D)} \leq c_D \|\varphi\|_{\tilde{H}^{-1}(D)}$ .

*Proof.* Since  $\tilde{H}^{-1}(D)$  can be regarded as the closure of  $C_0^\infty(D)$  with respect to  $\|\cdot\|_{H^{-1}(\mathbb{R}^d)}$ , we may consider  $\varphi \in C_0^\infty(D)$  and define

$$u(x) := (N_K \varphi)(x) = \int_K S(x, y) \varphi(y) \, dy.$$

Let  $\xi \in C_0^\infty([0, \infty))$  be a non-negative cut-off function satisfying  $\xi(t) = 1$  for  $t \in [0, 2R]$  and  $\xi(t) = 0$  for  $t > 3R$ , where  $R > 0$  is chosen such that  $D \subset B_R(0)$ . With

$$u_\xi(x) := \int_{\mathbb{R}^d} \xi(|x - y|) S(x, y) \varphi(y) \, dy.$$

it holds that  $u_\xi(x) = u(x)$  for  $x \in D$  and thus  $\|u\|_{H^1(D)} = \|u_\xi\|_{H^1(D)} \leq \|u_\xi\|_{H^1(\mathbb{R}^d)}$ . In [34, p. 109] it is proved that

$$\|u_\xi\|_{H^1(\mathbb{R}^d)} \leq c \|\varphi\|_{H^{-1}(\mathbb{R}^d)}$$

with  $c = c(R)$ . The assertion follows from

$$\|\varphi\|_{H^{-1}(\mathbb{R}^d)} = \sup_{0 \neq v \in H^1(\mathbb{R}^d)} \frac{(\varphi, v)_{L^2(\mathbb{R}^d)}}{\|v\|_{H^1(\mathbb{R}^d)}} \leq \sup_{0 \neq v \in H^1(D)} \frac{(\varphi, v)_{L^2(D)}}{\|v\|_{H^1(D)}} = \|\varphi\|_{\tilde{H}^{-1}(D)}.$$

□

**Lemma 8.** *Let  $u \in H^1(D)$  be harmonic. Then the norm of  $u$  in  $D$  is bounded by the norm in the  $C^{1,1}$  domain  $K \subset D$ , i.e.*

$$\|u\|_{1,D} \leq c_{D,K} \|u\|_{1,K}$$

with  $c_{D,K} > 0$  independent of  $u$ .

*Proof.* Let  $K^c := D \setminus \overline{K}$ . Since  $u$  is harmonic in  $D$ , it can be represented by its Cauchy data

$$u(x) = \begin{cases} (V\partial_\nu u)(x) - (Ku)(x), & x \in K, \\ -(V\partial_\nu u)(x) + (Ku)(x), & x \in K^c, \end{cases}$$

with the boundary integral operators

$$(Vu)(x) := \int_{\partial K} S(x, y)u(y) \, ds_y \quad \text{and} \quad (Ku)(x) := \int_{\partial K} \partial_{\nu, y} S(x, y)u(y) \, ds_y.$$

Since  $C_0^\infty(D)$  is dense in  $\tilde{H}^{-1}(D)$ , we may consider  $\varphi \in C_0^\infty(D)$  and estimate

$$\begin{aligned} \int_D u(x)\varphi(x) \, dx &= \int_K u(x)\varphi(x) \, dx + \int_{K^c} u(x)\varphi(x) \, dx \\ &= \int_{\partial K} (\partial_\nu u)(y)(N_K \varphi)(y) \, ds_y - \int_{\partial K} u(y)(\partial_\nu N_K \varphi)(y) \, ds_y \\ &\quad - \int_{\partial K} (\partial_\nu u)(y)(N_{K^c} \varphi)(y) \, ds_y + \int_{\partial K} u(y)(\partial_\nu N_{K^c} \varphi)(y) \, ds_y \\ &\leq \|\partial_\nu u\|_{H^{-1/2}(\partial K)} \|N_K \varphi\|_{H^{1/2}(\partial K)} + \|u\|_{H^{1/2}(\partial K)} \|\partial_\nu N_K \varphi\|_{H^{-1/2}(\partial K)} \\ &\quad + \|\partial_\nu u\|_{H^{-1/2}(\partial K)} \|N_{K^c} \varphi\|_{H^{1/2}(\partial K)} + \|u\|_{H^{1/2}(\partial K)} \|\partial_\nu N_{K^c} \varphi\|_{H^{-1/2}(\partial K)} \\ &\leq 2c'_K \|u\|_{H^1(K)} (\|N_K \varphi\|_{H^1(K)} + \|N_{K^c} \varphi\|_{H^1(K)}) \\ &\leq 2c'_K \|u\|_{H^1(K)} (\|N_K \varphi\|_{H^1(D)} + \|N_{K^c} \varphi\|_{H^1(D)}) \end{aligned}$$

due to  $\|u\|_{H^{1/2}(\partial K)} \leq \|u\|_{H^1(K)}$  and  $\|\partial_\nu u\|_{H^{-1/2}(\partial K)} \leq c'_K \|u\|_{H^1(K)}$ . Lemma 7 leads to

$$\int_D u(x)\varphi(x) \, dx \leq 4c'_K c_D \|u\|_{H^1(K)} \|\varphi\|_{\tilde{H}^{-1}(D)}$$

and thus

$$\|u\|_{H^1(D)} = \sup_{\varphi \in \tilde{H}^{-1}(D)} \frac{(u, \varphi)_{L^2(D)}}{\|\varphi\|_{\tilde{H}^{-1}(D)}} \leq 4c'_K c_D \|u\|_{H^1(K)},$$

which leads to the desired estimate  $\|u\|_{1,D} \leq c_{D,K} \|u\|_{1,K}$  with a constant  $c_{D,K} > 0$  depending on  $D$  and  $K$ .  $\square$

## References

- [1] J. Barnes and P. Hut. A hierarchical  $\mathcal{O}(N \log N)$  force calculation algorithm. *Nature*, 324:446–449, 1986.
- [2] M. Bebendorf. Efficient inversion of Galerkin matrices of general second-order elliptic differential operators with nonsmooth coefficients. *Math. Comp.*, 74:1179–1199, 2005.
- [3] M. Bebendorf. Why finite element discretizations can be factored by triangular hierarchical matrices. *SIAM J. Num. Anal.*, 45(4):1472–1494, 2007.

- [4] M. Bebendorf. *Hierarchical Matrices: A Means to Efficiently Solve Elliptic Boundary Value Problems*, volume 63 of *Lecture Notes in Computational Science and Engineering (LNCSE)*. Springer, 2008. ISBN 978-3-540-77146-3.
- [5] M. Bebendorf and T. Fischer. On the purely algebraic data-sparse approximation of the inverse and the triangular factors of sparse matrices. *Num. Lin. Alg. Appl.*, 18:105–122, 2011.
- [6] M. Bebendorf and W. Hackbusch. Existence of  $\mathcal{H}$ -matrix approximants to the inverse FE-matrix of elliptic operators with  $L^\infty$ -coefficients. *Numer. Math.*, 95(1):1–28, 2003.
- [7] M. Bebendorf and J. Ostrowski. Parallel hierarchical matrix preconditioners for the curl-curl operator. *Journal of Computational Mathematics, special issue on Adaptive and Multilevel Methods for Electromagnetics*, 27(5):624–641, 2009.
- [8] L. Berlyand and H. Owhadi. Flux norm approach to finite dimensional homogenization approximations with non-separated scales and high contrast. *Arch. Rational Mech. Anal.*, 198:677–721, 2010.
- [9] J. H. Bramble. *Multigrid methods*. Longman Scientific & Technical, 1993.
- [10] M. Dryja, M. V. Sarkis, and O. B. Widlund. Multilevel Schwarz methods for elliptic problems with discontinuous coefficients in three dimensions. *Numer. Math.*, 72(3):313–348, 1996.
- [11] W. E and B. Engquist. The heterogeneous multcale methods. *Comm. Math. Sci.*, 1(1):87–132, 2003.
- [12] Y. Efendiev and T.Y. Hou. *Multiscale Finite Element Methods: Theory and Applications*. Springer, 2009.
- [13] B. Engquist and H. Zhao. Approximate separability of Green’s function for high frequency Helmholtz equations. Technical Report CAM14-71, Department of Mathematics, University of California, Los Angeles, 2014.
- [14] C. Farhat, M. Lesoinne, P. Le Tallec, K. Pierson, and D. Rixen. FETI-DP: a dual-primal unified FETI method – part I. A faster alternative to the two-level FETI method. *International J. Numer. Methods Engrg.*, 50(7):1523–1544, 2001.
- [15] C. Farhat and F.-X. Roux. A method of finite element tearing and interconnecting and its parallel solution algorithm. *Int. J. Numer. Meth. Engrg.*, 32:1205–1227, 1991.
- [16] J. Galvis and Y. Efendiev. Domain decomposition preconditioners for multiscale flows in high contrast media: reduced dimension coarse spaces. *Multiscale Model. Simul.*, 8(5):1621–1644, 2010.
- [17] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Springer, Berlin, 2001. Reprint of the 1998 edition.
- [18] L. Grasedyck and W. Hackbusch. Construction and arithmetics of  $\mathcal{H}$ -matrices. *Computing*, 70:295–334, 2003.

- [19] L. F. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *J. Comput. Phys.*, 73(2):325–348, 1987.
- [20] L. F. Greengard and V. Rokhlin. A new version of the fast multipole method for the Laplace equation in three dimensions. In *Acta numerica, 1997*, volume 6 of *Acta Numer.*, pages 229–269. Cambridge Univ. Press, Cambridge, 1997.
- [21] M. Grüter and K.-O. Widman. The Green function for uniformly elliptic equations. *Manuscripta Math.*, 37(3):303–342, 1982.
- [22] W. Hackbusch. *Multi-Grid Methods and Applications*. Springer, 1985.
- [23] W. Hackbusch. A sparse matrix arithmetic based on  $\mathcal{H}$ -matrices. Part I: Introduction to  $\mathcal{H}$ -matrices. *Computing*, 62(2):89–108, 1999.
- [24] W. Hackbusch. *Hierarchische Matrizen*. Springer Verlag, 2009.
- [25] W. Hackbusch and B. N. Khoromskij. A sparse  $\mathcal{H}$ -matrix arithmetic. Part II: Application to multi-dimensional problems. *Computing*, 64(1):21–47, 2000.
- [26] T. Y. Hou and X. Wu. A multiscale finite element method for elliptic problems in composite materials and porous media. *J. Comput. Phys.*, 134:169–189, 1997.
- [27] A. Klawonn, O. B. Widlund, and M. Dryja. Dual-primal FETI methods for three-dimensional elliptic problems with heterogeneous coefficients. *SIAM J. Numer. Anal.*, 40(1):159–179, 2002.
- [28] N. Neuss, W. Jäger, and G. Wittum. Homogenization and multigrid. *Computing*, 66:1–26, 2001.
- [29] C. Pechstein, M. Sarkis, and R. Scheichl. New theoretical coefficient robustness results for FETI-DP. In R. Bank, R. Kornhuber, and O. Widlund, editors, *Domain Decomposition Methods in Science and Engineering XX*, Lecture Notes in Computational Science and Engineering. Springer-Verlag, 2013.
- [30] C. Pechstein and R. Scheichl. Scaling up through domain decomposition. *Appl. Anal.*, 88(10):1589–1608, 2009.
- [31] C. Pechstein and R. Scheichl. Weighted Poincaré inequalities and applications in domain decomposition. In Y. Huang, R. Kornhuber, O. Widlund, and J. Xu, editors, *Domain Decomposition Methods in Science and Engineering XIX*, pages 197–204, Berlin, 2010. Springer-Verlag.
- [32] C. Pechstein and R. Scheichl. Analysis of FETI methods for multiscale PDEs. *Numer. Math.*, 111(2):293–333, 2008.
- [33] J. W. Ruge and K. Stüben. Algebraic multigrid. In S. F. McCormick, editor, *Multigrid Methods*, pages 73–130. SIAM, 1987.
- [34] O. Steinbach. *Numerische Näherungsverfahren für elliptische Randwertprobleme: Finite Elemente und Randelemente*. B. G. Teubner, Stuttgart, Leipzig, Wiesbaden, 2003.
- [35] E. E. Tyrtysnikov. Mosaic-skeleton approximations. *Calcolo*, 33(1-2):47–57, 1996. Toeplitz matrices: structures, algorithms and applications (Cortona, 1996).